

Ground-truth dataset and baseline evaluations for base-detail separation algorithms at the part level

Xuan Dong*, Boyan I. Bonev*, Weixin Li✉, Weichao Qiu, Xianjie Chen, Alan L. Yuille, *Fellow, IEEE*,

Abstract—Base-detail separation is a fundamental image processing problem which models the image by a smooth base layer for the coarse structure, and a detail layer for the texture-like structures. Base-detail separation is hierarchical, and can be performed from the fine level to the coarse level. The separation at coarse level, in particular at the part level, is important for many applications, but currently lacks ground-truth datasets which are needed for comparing algorithms quantitatively. Thus, we propose a procedure to construct such datasets, and provide two examples: Pascal Part UCLA and Fashionista, containing 1000 and 250 images, respectively. Our assumption is that the base is piecewise smooth and we label the appearance of each piece by a polynomial model. The pieces are objects and parts of objects, obtained from human annotations. Finally, we propose a way to evaluate different separation methods with our datasets and compared the performances of seven state-of-the-art algorithms.

Index Terms—base-detail separation, part level.

I. INTRODUCTION

Base-detail separation is a fundamental problem in image processing, and is useful for a number of applications such as contrast enhancement [1], exposure correction [2], etc. It defines a simplified coarse representation of an image with its basic structures (base layer), and a detailed representation which may contain texture, fine details or just noise (detail layer), as shown in Fig 1.

This definition leaves open what is detail and what is base. We argue that base-detail separation should be formulated as a hierarchical structure. For instance, in an image of a crowd of people, we argue that their heads and faces form a texture, or detail, over a common base surface which could be their average color. At a less coarse level, we could say that each individual head is composed of two base regions: the hair and the face, where the details are the hair texture, the eyes, nose and mouth. We could go into still more detail and argue that the mouth of a person could also be separated into a smooth surface (or base), and the details of the lips, if there is enough resolution. Roughly speaking, from fine to coarse, the hierarchical base-detail separation can be classified as the pixel level, the sub-part level, the part level, and the object level.

Xuan Dong, Boyan I. Bonev, Weixin Li, Weichao Qiu, Xianjie Chen and Alan L. Yuille are with the Department of Statistics, UCLA, CA 90095. Alan L. Yuille is also with the Department of Cognitive Science and Computer Science, John Hopkins University, Baltimore, MD 21218. (e-mail: dongx@ucla.edu; boyan@ua.es; lwx@cs.ucla.edu; qiuwch@gmail.com; cxj@ucla.edu; alan.l.yuille@gmail.com) Corresponding author: Weixin Li.

*Xuan Dong and Boyan I. Bonev contributed to this work equally.

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

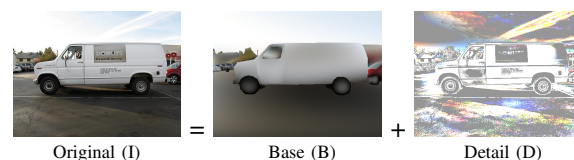


Fig. 1. Base-Detail separation is a fundamental image processing problem. It relates to a wide range of tasks, such as contrast enhancement, exposure correction, etc. (Here, D is shown with increased contrast for clarity).

Fig. 2 shows an example of hierarchical base-detail separation of an image.

Benchmark datasets are becoming important in computer vision and image processing. For most computer vision problems, like optical flow [3], stereo [4], object recognition [5], and edge detection [6], there exist datasets used as benchmarks for evaluation and comparison. These datasets have driven innovation and rigor to those problems. However, there is a lack of a common dataset for image base-detail separation at coarser levels of the hierarchy, in particular, at the part level. This makes it hard to draw conclusions and to compare quantitatively.

Base-detail separation at the part level is important for many applications, and failing to do it correctly will introduce artifacts into the final enhancement results. Fig. 4 shows the types of errors, such as halo artifacts, in exposure correction [2], which result from incorrect separation. Some examples of the incorrect separation are shown in Fig. 3. With these figures we do not intend to show best exposure-correction quality for human perception. We are showing the halo artifacts when using uncorrect separation results, comparing with the halo-free enhancement of the proposed ground-truth.

Thus, we generate a ground-truth base-detail dataset at the part level in this paper. At the part level, a fundamental assumption of this paper is that the base layers are piecewise smooth. It is piecewise because of the different objects and parts present in the image. For each part at the part level, the separation results should not be affected by their neighboring parts, hence the methods should successfully preserve the sharp boundaries. Otherwise, halo artifacts will be introduced into both of the base and the detail, as shown in Fig. 3.

To get the ground-truth data, we manually segment each image into parts, and for each part, we label the base and detail layers. Segmenting images into parts is challenging because the shapes of parts and their appearances vary a lot. We rely on manual annotations of the segments at the pixel level. Within each part, labeling the base-detail separation is also challenging because it requires pixel-wise annotation for the intensity of the base layer in the RGB color space (or, more

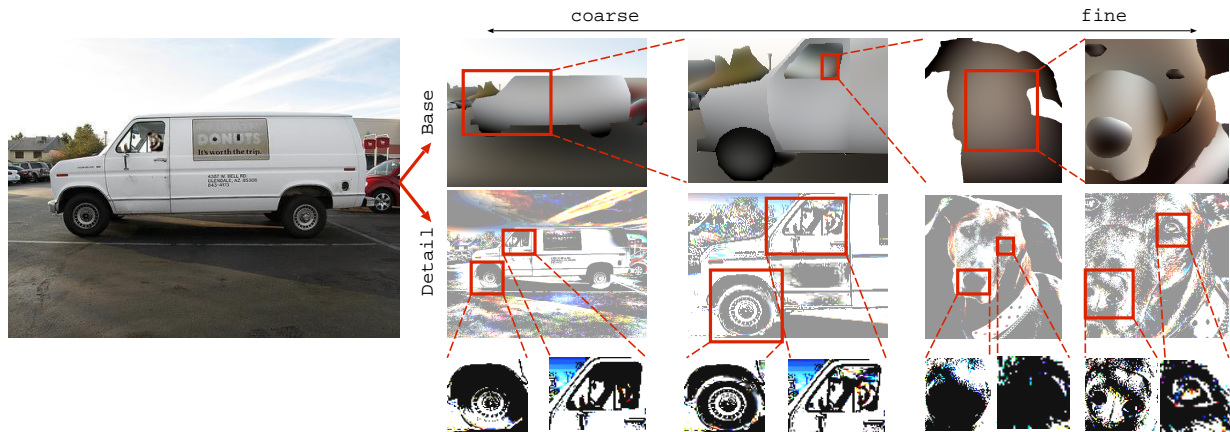


Fig. 2. Example of hierarchical base-detail separation. The original image (left) is decomposed into base (first row) and detail (second row). The columns are sorted from coarser to finer taxonomy in the hierarchy. Depending on the hierarchy level, details contain different information, e.g., the nose of the dog can be considered as detail, or it can be considered as a part and the nostrils are the detail (see third row). Note that detail is represented with a very high contrast here, for better visualization.

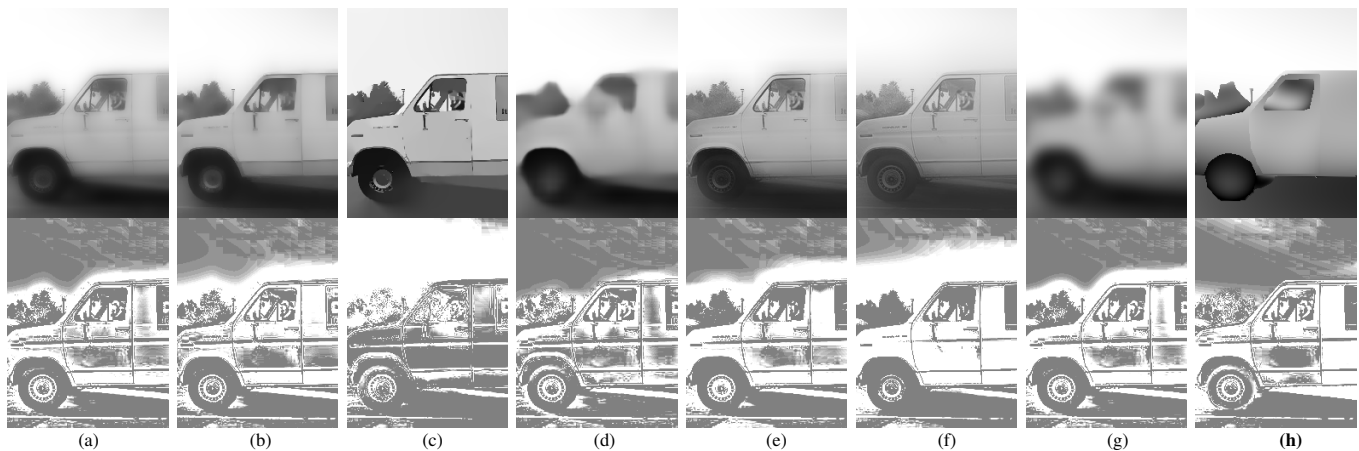


Fig. 3. The halo effects produced by different filters, (a) to (g), compared to the proposed ground-truth (h). Top row: base. Bottom row: detail, with increased contrast for visualization. Filters: (a) adaptive manifold (AM) [8], (b) domain transform (DT) [7], (c) L0 smooth (L0) [1], (d) rolling guidance (RG) [9], (e) bilateral (BL) [10], (f) guided (GD) [11], and (g) Gaussian (GS). The different halo effects are easily observed between the sky and the vehicle, and most of them produce blurry boundaries in the base layer and a padding artifact in the detail layer. Another effect is that some methods represent the cloud as a broken texture (detail layer), and not as a blob-like structure (groundtruth detail, (h)).

generally, in any color space). We use a polynomial model of different orders to separate the image signal into several possible base layers and let humans select which order of the polynomial separation is the correct base layer separation. The residual of the selected base layer is the detail. It is possible that none of the polynomial model's results are correct for the base layer. So, in the annotation, we exclude the images if even one region of the image cannot be described by the polynomial model.

The main contributions of this paper are: 1) two natural image datasets providing the part level base and detail ground truth (Pascal Part UCLA and Fashionista); and 2) the evaluation of several state-of-the-art base-detail separation algorithms based on the provided ground truth. The supplementary materials include more detailed information of the related work and our work, and more experimental results.

II. RELATED WORK

A. Image base-detail separation algorithms and applications

We argue that the separation of image base-detail layers is hierarchical. And different image base-detail separation algorithms focus on separation at different levels of the hierarchical structure, aiming at different applications.

Base-detail separation at fine level usually aims at the separation of signal and noise. The detail layer is occupied by noise while the base layer contains the image signal. Different algorithms have been proposed for applications such as denoising [12], joint upsampling [13], compression artifact removal [1], etc.

Base-detail separation at coarse level aims to separate the image's basic coarse structures, such as the luminance and color of parts, and texture-like detail, such as high frequency and mid-high frequency information. Different algorithms are proposed for different applications such as exposure correction [2], HDR/tone mapping [14], style transfer [15], tone manage-

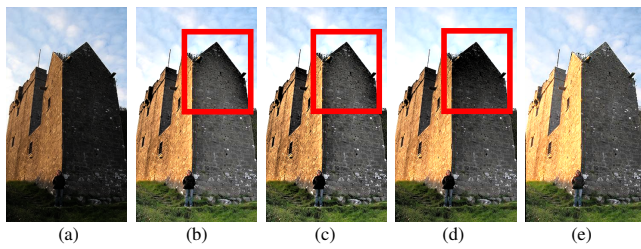


Fig. 4. Halo effects produced by different filters for exposure correction of input image (a). (b): The result of the adaptive manifold filter. (c): The result of the domain transform filter. (d): The result of the guided filter. (e): The result using the proposed ground-truth base-detail separation result. The halo effects are easily observed between the sky and the building (as marked in the red box region), which are caused by incorrect base-detail separation at the part level. Please see more examples in the supplementary material.

ment [16], contrast enhancement [7], image abstraction [1], etc.

B. Related datasets and quantitative evaluation

There are some quantitative evaluation methods for base-detail separation at the fine level, suitable for image denoising [12], upsampling [13], and compression artifact removal [1]. But little work have been proposed for quantitative evaluation at the part level. The performance of base-detail at the fine level has little relationship with the performance at the part level. This is because ignoring the piecewise smoothness assumption does not have big effects on the results at fine level, but can have a very big effect at the part level. For example, at the part level, artifacts like halos are frequently introduced because the separation of a part is contaminated by neighboring parts. Thus, ground truth base-detail separation at the part level is lacked and desired.

III. THE GROUND-TRUTH BASE-DETAIL SEPARATION DATASET

The goal of this paper is to construct a ground-truth base-detail separation dataset at the part level. We assume that the separated base layer is piecewise smooth (validated by our human annotators). It means that the separation of pixels of one part is only determined by the part itself, and the neighboring parts do not affect it. This avoids the halo effects between parts (see Fig. 4 and 3). Thus, to get the ground-truth base-detail separation dataset, 1) we proposed to use images that are manually segmented at the part level, and 2) within each part, we annotate the base layer in the RGB color space by using a polynomial model to separate each part into several possible base layers and letting humans select the correct one.

We rely on pixel-wise human annotations for the part level segmentation because the shapes of different parts vary a lot and pixel-wise annotations could get accurate segmentation results. There exists many automatic segmentation algorithms, e.g. [17], but they are not accurate enough for our work. For base-detail separation within each part, we use both polynomial and human annotation because it is difficult for humans to directly annotate the base and detail layers of every pixel in the RGB color space. Thus, we proposed to first separate the image signals of each part into several possible results. Then, from the set of possible separation results, we let

humans select which one gives the best base-detail separation of each part. We use a polynomial model with different orders, which is one of the most basic signal processing methods, to produce several possible results for each part. This strategy reduces the labeling work a lot and makes it much easier to generate the ground-truth datasets for base-detail separation at the part level.

A. Annotation for segmentation

The annotation for segmentation at the part level consists of drawing the exact boundaries of each part in the image. Fortunately, there exist some well known datasets with this kind of annotations, like the Fashionista [18] and the Pascal Part UCLA dataset [19]. They provide human-made annotations of parts at the pixel level. We reuse their part-level annotations. In Fashionista, the labeled regions are parts of human beings, such as hair, glasses, etc., and the segmented parts have appearance consistency. In the Pascal Part UCLA dataset, the labeled images have many classes, such as dog, person, etc. And regions are labeled according to semantic meaning and do not necessarily enforce the appearance consistency.

The definition of a part is different in different applications and datasets because the base-detail separation is a hierarchical structure. This explains why the Fashionista and the Pascal Part UCLA datasets used different strategies for parts annotation. In our opinion, this diversity of the labeling of parts is a good property because we can evaluate different base-detail separation algorithms at different levels of the hierarchical structure so that we can have a better understanding of the performance of different algorithms.

B. Annotation for base-detail separation

To annotate the base layer within each part, the first step is to separate the image signals of each part into several possible base layers with the polynomial model at different orders. Within each part P_i , we fit polynomial models on each color channel separately. The number of parameters $\vec{\omega}$ depends on the order k of the polynomial. The polynomial approximations are :

$$\begin{aligned} b_k(\vec{x}, \vec{\omega}) &= \vec{x}^T \vec{\omega} \\ k=0: \vec{x} &= 1, \vec{\omega} = \omega_0 \\ k=1: \vec{x} &= [1, x_1, x_2], \vec{\omega} = [\omega_0, \omega_1, \omega_2] \\ k=2: \vec{x} &= [1, x_1, x_2, x_1^2, x_2^2, x_1x_2], \vec{\omega} = [\omega_0, \dots, \omega_5] \\ k=3: \vec{x} &= [1, x_1, x_2, x_1^2, x_2^2, x_1x_2, x_1^3, x_2^3, x_1x_2^2, \\ &\quad x_1^2x_2], \vec{\omega} = [\omega_0, \dots, \omega_9] \end{aligned} \quad (1)$$

The estimation of the parameters $\vec{\omega}$ of the polynomial is performed by linear least squares QR factorization [20]. We limit our polynomial approximations to the third order, $k=3$, to prevent over-fitting the data.

The second step is to select the ground-truth base layer for each part. For each part, there are four possible ground-truth base layers, and we let the annotators select the ground-truth base layer by choosing one from the four layers.

After the previous two steps, we get the ground-truth base layer using the polynomial model's separation results and the

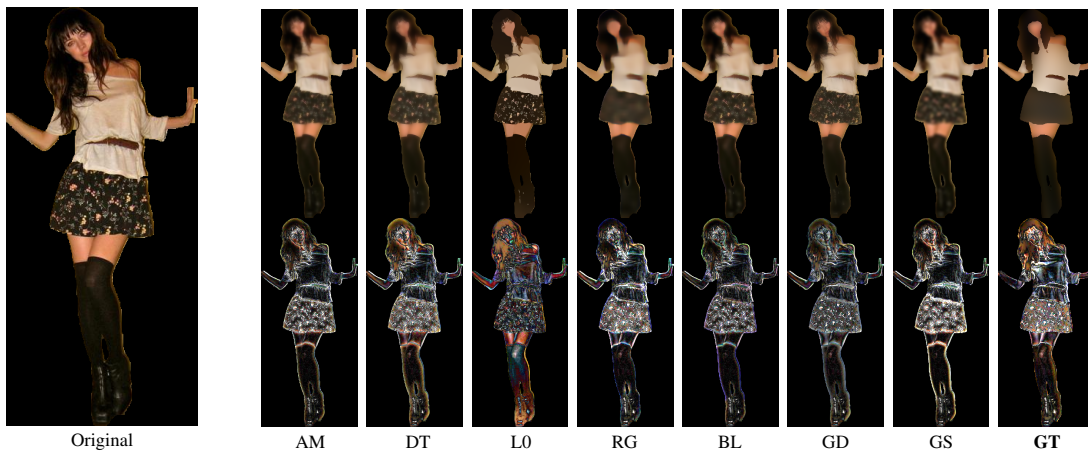


Fig. 5. Results of the base-detail separation methods tested on an example image of the Fashionista dataset. Top: base; Bottom: detail. Last column (GT) is the proposed ground-truth.

annotation of the annotators on each part. It is possible that none of the possible results obtained by the polynomial model are correct for the base layer. Thus, for an image, if the base layer of even one part cannot be described by any of the polynomial results, this image is rejected. In this way, we get a subset of the images from the whole datasets of the Fashionista and the Pascal Part UCLA, for which the base layer can be described by some order of the polynomial model. In total, we select about 1000 and 250 images in the Pascal Part UCLA and the Fashionista datasets, respectively.

In our labeling, 15 annotators in total performed the labeling separately. So, for each region of the images, we have 15 annotations. If 7 or more of the 15 annotations choose ‘outlier region’, we will see this region as an outlier to be modeled by the polynomial model and do not select the image of the object into our final datasets. Otherwise, the order of the polynomial for this region is voted by the 15 annotations. And the base layer of the region is reconstructed by the polynomial model.

IV. EVALUATION

A. Ground-truth datasets

The ground-truth datasets that we use are subsets of images from the Fashionista [18] and the Pascal Part UCLA datasets [19], as described in Sec. III. For simplification, in this paper, we still call the subset of images the Fashionista and the Pascal Part UCLA datasets, respectively. See examples of both datasets in Fig. 6.

B. Separation methods

The separation methods we use in the evaluation include the adaptive manifold filter (AM) [8], the domain transform filter (DT) [7], the L0 smooth filter (L0) [1], the rolling guidance filter (RG) [9], the bilateral filter (BL) [10], the guided filter (GD) [11], the Gaussian filter (GS). The Gaussian filter is a linear filter, and the smoothing only considers the distance between neighboring pixels. The other filters are edge-preserving filters.

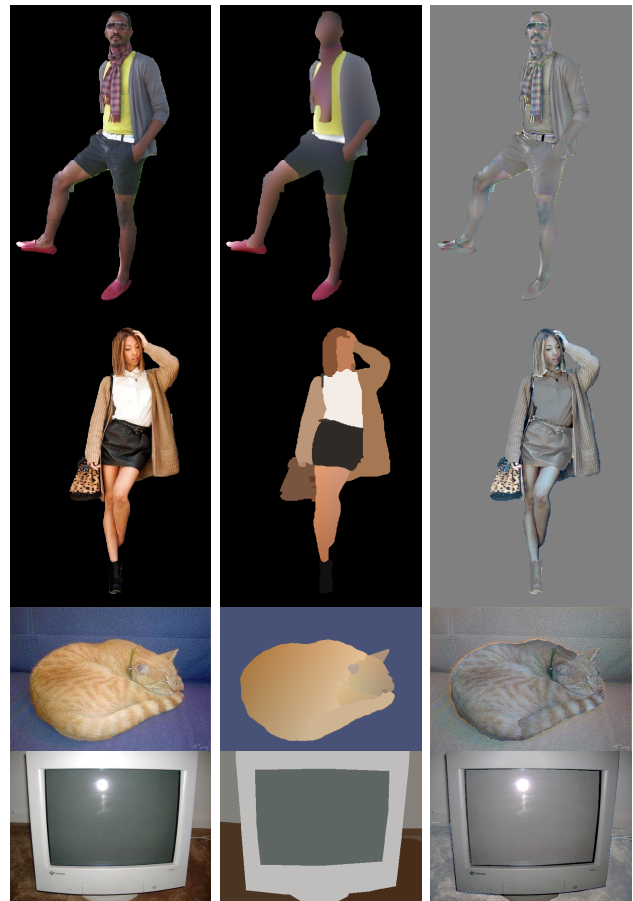


Fig. 6. Examples of the images in the Fashionista (the top two) and the Pascal Part UCLA (the bottom two) base-detail datasets. From left to right: original image, base and detail.

C. Error metric

A direct way for evaluation is to compute the mean squared error (MSE) between ground-truth base-detail layers and estimated base-detail layers. MSE is defined as:

$$MSE(J_1, J_2) = \frac{1}{\sum_{i,c} 1} \sum_{i,c} (J_1(i, c) - J_2(i, c))^2,$$

TABLE I
PARAMETERS SETTINGS OF DIFFERENT FILTERS FOR THE FASHIONISTA AND THE PASCAL PART UCLA DATASETS. **LEFT:** THE PARAMETERS RANGE OF THE FILTERS. **RIGHT:** THE OPTIMAL PARAMETERS OF THE METHODS FOR THE DATASETS.

Alg.	Ranges of parameters θ_i		Optimal (θ_1, θ_2)	
	θ_1	θ_2	Fashionista	Pascal
AM	2,4,8,16,32,64	0.1,0.2,0.4,0.8,1,2	(8,0.8)	(32,0.8)
DT	4,8,16,32,64,128	0.1,0.2,0.4,0.8,1,2	(16,2)	(64,2)
L0	0.005,0.01,0.02,0.04,0.08,0.1,0.2,0.4,0.8,1		0.08	0.1
RG	2,4,8,16	0.04,0.08,0.1,0.2,0.4	(8,0.1)	(16,0.2)
BL	2,4,8,16,32,64	0.1,0.2,0.4,1,2,4	(8,0.4)	(32,0.4)
GD	4,8,16,32,64,128	0.01,0.04,0.1,0.2,0.4	(16,0.1)	(64,0.1)
GS	4,8,16,32,64,128	1,2,4,8,16,32	(32,4)	(64,4)

where J_1 and J_2 are two images, i is the pixel position, c is the color channel in the RGB space.

However, we found that the same amount of error will cause very different MSE values for well-exposed images and low lighting images. For example, the intensities of a pixel in a well-exposed image and low lighting image are 200 and 20, respectively. If the errors are the same, for example 10%, the MSE values will be very different (400 and 4, respectively). The reason is for well-exposed images, because the RGB intensities of pixels are high, small errors will lead to large MSE values. So directly using MSE will lead to bias to the evaluation results and the errors of well-exposed images will have more weights. To reduce the bias, we proposed the relative mean squared error (RMSE) as the error metric between the ground-truth base-detail layers and the estimated base-detail layers. RMSE is defined by

$$RMSE(B_{GT}, D_{GT}, B_E, D_E) = \frac{1}{2} \left(\frac{MSE(B_{GT}, B_E)}{MSE(B_{GT}, 0)} + \frac{MSE(D_{GT}, D_E)}{MSE(D_{GT}, 0)} \right), \quad (2)$$

where, B_{GT} is the ground truth base layer, D_{GT} is the ground-truth detail layer, B_E is the estimated base layer, and D_E is the estimated detail layer. Because 1) RMSE considers errors of both the detail and the base layers, and 2) for each layer, it measures the relative error, i.e. the ratio of $MSE(GT, E)$ and $MSE(GT, 0)$, it reduces the bias between low lighting images and well exposed images. According to the definition of RMSE, if the RMSE value is lower, the estimation of base and detail layers are more accurate.

D. Algorithms and parameter settings

For an input image, we use different algorithms to smooth it to obtain the base layers. Then, we compute the RMSE between the filtered result and the ground-truth data. The seven separation methods shown in Table I have parameters to control the smoothing. In general, high values of the parameters tend to mean coarse level smoothing. Here, we select the best parameters for each filter to enable a fair comparison between them. We use the same parameter value for the whole dataset (one parameter for each dataset). The parameters range of the filters and the optimal parameters are shown in Table I. For GS and GD, θ_1, θ_2 denote window size and σ variance, respectively. For BL, AM, DT, and RG, θ_1, θ_2 denote σ spatial and σ range, respectively. For L0, θ_1 denotes λ . The parameters for the Fashionista and the Pascal Part UCLA datasets are different, because, as described

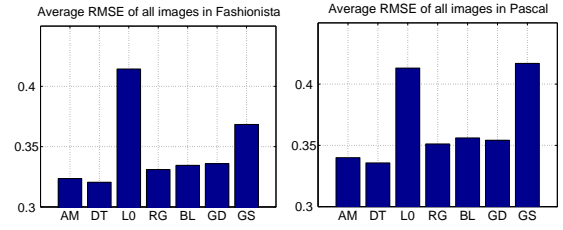


Fig. 7. Quantitative comparison of the seven separation algorithms on the datasets (from left to right): average RMSE on the Fashionista and the Pascal Part UCLA (over all images in the dataset).

in Sec. III-A, these two datasets used different strategies for part annotation. The segmented parts of the Pascal Part UCLA dataset are usually coarser than those of the Fashionista dataset. As a result, the optimal parameters for the Pascal Part UCLA dataset are usually larger than those for the Fashionista dataset. The results are shown in Fig. 7. Some example results of each method can be seen in Figs. 3, 5.

E. Analysis

We can see that most of the edge-preserving filters perform better than the Gaussian filter (we consider the Gaussian filter to be the classical baseline). It's because the Gaussian filter does not preserve edges, and the high parameters (e.g., the variance) lead to the smoothing results of each part affected heavily by neighboring parts. Most of the other filters preserve edges better than the Gaussian filter, so they have better performance. The bilateral filter is edge-preserving and consistent with standard intuition, so it performs better than the Gaussian filter. The adaptive manifold filter and domain transform filter have better performance than the bilateral filter on average because they are flexible, and have more potential to perform well in our datasets if the parameters are selected carefully. The rolling guidance filter and guided filter also have good performances in the experiments. The L0 smooth filter makes use of gradient to separate base and detail. However, in our datasets, the parts are segmented semantically, and areas with large gradient do not always mean the edges of parts. This results in poor performance.

V. CONCLUSIONS

Quantitative evaluations are fundamental to the advances of any research field. The part level base-detail ground-truth datasets we provide here are a necessary starting point for extensive quantitative comparisons of the base-detail separation algorithms at the part level. We argue that, ideally, base-detail annotation should be hierarchical, and we proposed an intermediate solution, which is practical (i.e., ready for use) by now and extensible in the future.

VI. ACKNOWLEDGEMENT

We would like to thank the anonymous reviewers for their help in improving the paper. This work is partially supported by NSF CCF-1317376.

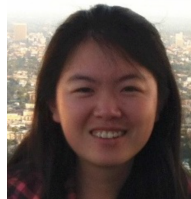
REFERENCES

- [1] L. Xu, C. Lu, Y. Xu and J. Jia, "Image smoothing via L0 gradient minimization", *TOG*, vol. 30, no. 6, pp. 1–14, 2011.
- [2] L. Yuan and J. Sun, "Automatic exposure correction of consumer photographs", *ECCV*, pp. 771–785, 2012.
- [3] D. J. Butler, J. Wulff, G. B. Stanley and M. J. Black, "A naturalistic open source movie for optical flow evaluation", *ECCV*, pp. 611–625, 2012.
- [4] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *IJCV*, vol. 47, no. 1–3, pp. 7–42, 2002.
- [5] L. Fei-Fei, R. Fergus and P. Perona, "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories", *CVIU*, vol. 106, no. 1, pp. 59–70, 2007.
- [6] P. Arbelaez, M. Maire, C. Fowlkes and J. Malik, "Contour detection and hierarchical image segmentation", *PAMI*, vol. 33, no. 5, pp. 898–916, 2011.
- [7] E. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing", *TOG*, vol. 30, no. 4, pp. 69, 2011.
- [8] E. Gastal and M. Oliveira, "Adaptive manifolds for real-time high-dimensional filtering", *TOG*, vol. 31, no. 4, pp. 33, 2012.
- [9] Q. Zhang, X. Shen, L. Xu and J. Jia, "Rolling guidance filter", *ECCV*, pp. 815–830, 2014.
- [10] S. Paris, P. Kornprobst, J. Tumblin and F. Durand, "Bilateral filtering: theory and applications", *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 1, pp. 1–73, 2008.
- [11] K. He, J. Sun and X. Tang, "Guided image filtering", *ECCV*, pp. 1–14, 2010.
- [12] T. Brox and D. Cremers, "Iterated nonlocal means for texture restoration", *International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 13–24, 2007.
- [13] P. Bhat, C. L. Zitnick, M. Cohen and B. Curless, "GradientShop: A gradient domain optimization framework for image and video filtering", *TOG*, vol. 29, no. 2, pp. 10, 2010.
- [14] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high dynamic range images", *TOG*, vol. 21, no. 3, pp. 257–266, 2002.
- [15] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz and F. Durand, "Fast local laplacian filters: theory and applications", *TOG*, vol. 33, no. 5, pp. 167, 2014.
- [16] S. Bae, S. Paris and F. Durand, "Two-scale tone management for photographic look", *TOG*, vol. 25, no. 3, pp. 637–645, 2006.
- [17] B. Bonev and A. L. Yuille, "A fast and simple algorithm for producing candidate regions", *ECCV*, 2014.
- [18] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz and T. L. Berg, "Parsing clothing in fashion photographs", *CVPR*, pp. 3570–3577, 2012.
- [19] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun and A. L. Yuille, "Detect what you can: detecting and representing objects using holistic models and body parts", *CVPR*, pp. 1971–1978, 2014.
- [20] G. H. Golub and C. Loan, "Matrix computations", *JHU Press*, vol. 3, 2012.



design of biometric algorithms and sensors for Synaptics Inc.

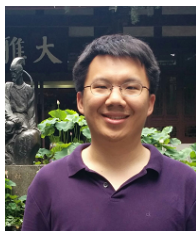
Boyan Bonev received a Ph.D. in Computer Science in 2010 from the University of Alicante, Spain, where he has been lecturer in the Economics Department and researcher in the Computer Science and Artificial Intelligence Department. His research is in the context of Mobile Robotics, Computer Vision and Natural Language Processing. He did postdoctoral research with Prof. Alan Yuille in the Statistics Department of the University of California, Los Angeles, working on image segmentation and object recognition. He is currently working on the



Weixin Li is currently pursuing the Ph.D. degree in Computer Science with the Center for Vision, Cognition, Learning and Autonomy (VCLA) at the University of California, Los Angeles (UCLA). She received the MS degree in Computer Science from UCLA in 2014, and the B.E. degree in Computer Science from Beihang University in 2010. Her current research interests include computer vision and big data analytics, with a focus on the study of news and social media.



Weichao Qiu received the B.E. and M.S. degree from Huazhong University of Science and Technology, Wuhan, China, in 2011 and 2014 respectively. He is currently a Ph.D. student in the computer science department of Johns Hopkins University. He transferred from UCLA to Johns Hopkins University in 2016 with Prof. Alan Yuille, before that he was a Ph.D. student of UCLA. His research interest is computer vision.



Xuan Dong received the Ph.D. degree in Computer Science from Tsinghua University in 2015, and the B.E. degree in Computer Science from Beihang University in 2010. He is currently a postdoctoral researcher with Prof. Alan Yuille in the Statistics Department of the University of California, Los Angeles. His research interests include computer vision and computational photography.



Xianjie Chen received the BS degree in computer science and technology with honor from Tsinghua University in 2011, and his PhD degree from UCLA in 2016. He is currently a research scientist at Facebook. His research interests include computer vision and machine learning with focus on object detection, human pose estimation and semantic image segmentation.



Alan L. Yuille received his B.A. in mathematics from the University of Cambridge in 1976, and completed his Ph.D. in theoretical physics at Cambridge in 1980 studying under Stephen Hawking. Following this, he held a postdoc position with the Physics Department, University of Texas at Austin, and the Institute for Theoretical Physics, Santa Barbara. He then joined the Artificial Intelligence Laboratory at MIT (1982-1986), and followed this with a faculty position in the Division of Applied Sciences at Harvard (1986-1995), rising to the position of associate professor. From 1995-2002 Alan worked as a senior scientist at the SmithKettlewell Eye Research Institute in San Francisco. In 2002 he accepted a position as full professor in the Department of Statistics at the University of California, Los Angeles. In January 2016 he joined Johns Hopkins University as a Bloomberg Distinguished Professor in Cognitive Science and Computer Science. He has over two hundred peer-reviewed publications in vision, neural networks, and physics, and has co-authored two books: *Data Fusion for Sensory Information Processing Systems* (with J. J. Clark) and *Two- and Three- Dimensional Patterns of the Face* (with P. W. Hallinan, G. G. Gordon, P. J. Giblin and D. B. Mumford); he also co-edited the book *Active Vision* (with A. Blake). He has won several academic prizes and is a Fellow of IEEE.